Zero-Gradient-Sum Algorithms for Distributed Convex Optimization: The Continuous-Time Case*

Jie Lu and Choon Yik Tang

School of Electrical and Computer Engineering

University of Oklahoma, Norman, OK 73019, USA

{jie.lu-1,cytang}@ou.edu

January 20, 2013

Abstract

This paper presents a set of continuous-time distributed algorithms that solve unconstrained, separable, convex optimization problems over undirected networks with fixed topologies. The algorithms are developed using a Lyapunov function candidate that exploits convexity, and are called *Zero-Gradient-Sum* (ZGS) algorithms as they yield nonlinear networked dynamical systems that evolve invariantly on a zero-gradient-sum manifold and converge asymptotically to the unknown optimizer. We also describe a systematic way to construct ZGS algorithms, show that a subset of them actually converge exponentially, and obtain lower and upper bounds on their convergence rates in terms of the network topologies, problem characteristics, and algorithm parameters, including the algebraic connectivity, Laplacian spectral radius, and function curvatures. The findings of this paper may be regarded as a natural generalization of several well-known algorithms and results for distributed consensus, to distributed convex optimization.

1 Introduction

This paper addresses the problem of solving an unconstrained, separable, convex optimization problem over an N-node multi-hop network, where each node i observes a convex function f_i, and all the N nodes wish to determine an optimizer x^* that minimizes the sum of the f_i's, i.e.,

$$x^* \in \arg \min_x \sum_{i=1}^{N} f_i(x). \tag{1}$$

The problem (1) arises in many emerging and future applications of multi-agent systems and wired/wireless/social networks, where agents or nodes often need to collaborate in order to jointly accomplish sophisticated tasks in decentralized and optimal fashions [1].

To date, a family of discrete-time *subgradient algorithms*, aimed at solving problem (1) under general convexity assumptions, have been reported in the literature. These subgradient algorithms may be roughly classified into two groups. The first group of algorithms [1–4] are *incremental* in nature, relying on the passing of an estimate of x^* around the network to operate. The second group of algorithms [5–7] are *non-incremental*, relying instead on a combination of subgradient updates and linear consensus iterations to operate, although gossip-based updates have also been considered [8]. For each of these algorithms, a number of convergence properties have been established, including the resulting error bounds, asymptotic convergence, and convergence rates.

In [9], we introduced two gossip-style, distributed asynchronous algorithms, referred to as *Pairwise Equalizing* (PE) and *Pairwise Bisectioning* (PB), which solve the scalar version of problem (1), in a manner that is fundamentally different from the aforementioned subgradient algorithms (e.g., PE and PB do not try to move along the gradient, nor do they require the notion of a stepsize). In [10], we showed that the two basic ideas behind PE—namely, the conservation of a certain gradient sum at zero and the use of a convexity-inspired Lyapunov function—can be extended, leading to *Controlled Hopwise Equalizing* (CHE), a distributed asynchronous algorithm that allows individual nodes to use potential drops in the value of the Lyapunov function to control, on their own, when to initiate an iteration, so that problem (1) may be solved efficiently. In both the papers [9, 10], problem (1) was studied in a discrete-time, asynchronous setting, and only the scalar version of it was considered.

In this paper, we address problem (1) from a continuous-time and multi-dimensional standpoint, building upon the two basic ideas behind PE. Specifically, assuming that each f_i in (1) is twice continuously differentiable and strongly convex and using the same Lyapunov function candidate as the one for PE and CHE, we first derive a family of continuous-time distributed algorithms called *Zero-Gradient-Sum* (ZGS) algorithms, with which the states of the resulting nonlinear networked dynamical systems slide along an invariant, zero-gradient-sum manifold and converge asymptotically to the unknown minimizer x^* in (1). We then describe a systematic way to construct ZGS algorithms and prove that a subset of them are exponentially convergent. For this subset of algorithms, we also obtain lower and upper bounds on their convergence rates as functions of the network topologies, problem characteristics, and algorithm parameters, including the algebraic connectivity, Laplacian spectral radius, and curvatures of the f_i's. As another contribution of this paper, we show that some of the existing continuous-time distributed consensus algorithms (e.g., [11–16]) are special cases of ZGS algorithms and are, interestingly, just a slight modification away from solving any problem of the form (1). In addition, the well-known result from [12], which says that the convergence rate of a linear consensus algorithm is characterized by the algebraic connectivity of the underlying graph, is a special case of Theorem 2 here.

2 Preliminaries

A twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is *locally strongly convex* if for any convex and compact set $D \subset \mathbb{R}^n$, there exists a constant $\theta > 0$ such that the following equivalent conditions hold [17, 18]:

$$f(y) - f(x) - \nabla f(x)^T (y - x) \geq \frac{\theta}{2} \|y - x\|^2, \quad \forall x, y \in D, \tag{2}$$

$$(\nabla f(y) - \nabla f(x))^T (y - x) \geq \theta \|y - x\|^2, \quad \forall x, y \in D, \tag{3}$$

$$\nabla^2 f(x) \geq \theta I_n, \quad \forall x \in D, \tag{4}$$

where $\| \cdot \|$ denotes the Euclidean norm, $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is the gradient of f, $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ is the Hessian of f, $I_n \in \mathbb{R}^{n \times n}$ is the identity matrix, and $\geq$ denotes matrix inequality (i.e., $A \geq B$ means $A - B$ is a positive semidefinite matrix). The function f is *strongly convex* if there exists a constant $\theta > 0$ such that the equivalent conditions (2)–(4) hold for $D = \mathbb{R}^n$, in which case θ is called the *convexity parameter* of f [18]. Finally, for any twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, any convex set $D \subset \mathbb{R}^n$, and any constant $\Theta > 0$, the following conditions are equivalent [18, 19]:

$$f(y) - f(x) - \nabla f(x)^T (y - x) \leq \frac{\Theta}{2} \|y - x\|^2, \quad \forall x, y \in D, \tag{5}$$

$$(\nabla f(y) - \nabla f(x))^T (y - x) \leq \Theta \|y - x\|^2, \quad \forall x, y \in D, \tag{6}$$

$$\nabla^2 f(x) \leq \Theta I_n, \quad \forall x \in D. \tag{7}$$

3 Problem Formulation

Consider a multi-hop network consisting of $N \geq 2$ nodes, connected by bidirectional links in a fixed topology. The network is modeled as a connected, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ represents the set of N nodes and $\mathcal{E} \subset \{\{i, j\} : i, j \in \mathcal{V}, i \neq j\}$ represents the set of links. Any two nodes $i, j \in \mathcal{V}$ are one-hop neighbors and can communicate if and only if $\{i, j\} \in \mathcal{E}$. The set of one-hop neighbors of each node $i \in \mathcal{V}$ is denoted as $\mathcal{N}_i = \{j \in \mathcal{V} : \{i, j\} \in \mathcal{E}\}$, and the communications are assumed to be delay- and error-free, with no quantization.

Suppose each node $i \in \mathcal{V}$ observes a function $f_i : \mathbb{R}^n \to \mathbb{R}$ satisfying the following assumption:

Assumption 1. For each $i \in \mathcal{V}$, the function f_i is twice continuously differentiable, strongly convex with convexity parameter $\theta_i > 0$, and has a locally Lipschitz Hessian $\nabla^2 f_i$.

Suppose, upon observing the f_i's, all the N nodes wish to solve the following unconstrained, separable, convex optimization problem:

$$\min_{x \in \mathbb{R}^n} F(x), \tag{8}$$

where the objective function $F : \mathbb{R}^n \to \mathbb{R}$ is defined as $F(x) = \sum_{i \in \mathcal{V}} f_i(x)$. The proposition below shows that F has a unique minimizer $x^* \in \mathbb{R}^n$, so that problem (8) is well-posed:

Proposition 1. *With Assumption 1, there exists a unique $x^* \in \mathbb{R}^n$ such that $F(x^*) \leq F(x)$ $\forall x \in \mathbb{R}^n$ and $\nabla F(x^*) = 0$.*

Proof. See Theorem 6 in [20]. □

Given the above network and problem, the aim of this paper is to devise a continuous-time distributed algorithm of the form

$$\dot{x}_i(t) = \varphi_i(x_i(t), \mathbf{x}_{\mathcal{N}_i}(t); f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall t \geq 0, \ \forall i \in \mathcal{V}, \tag{9}$$

$$x_i(0) = \chi_i(f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall i \in \mathcal{V}, \tag{10}$$

where $t \geq 0$ denotes time; $x_i(t) \in \mathbb{R}^n$ is a state representing node i's estimate of the unknown minimizer x^* at time t; $\mathbf{x}_{\mathcal{N}_i}(t) = (x_j(t))_{j \in \mathcal{N}_i} \in \mathbb{R}^{n|\mathcal{N}_i|}$ is a vector obtained by stacking $x_j(t)$ $\forall j \in \mathcal{N}_i$; $\mathbf{f}_{\mathcal{N}_i} = (f_j)_{j \in \mathcal{N}_i} : \mathbb{R}^n \to \mathbb{R}^{|\mathcal{N}_i|}$ is a function obtained by stacking f_j $\forall j \in \mathcal{N}_i$; $\varphi_i : \mathbb{R}^n \times \mathbb{R}^{n|\mathcal{N}_i|} \to \mathbb{R}^n$ is a locally Lipschitz function of $x_i(t)$ and $\mathbf{x}_{\mathcal{N}_i}(t)$ governing the dynamics of $x_i(t)$, whose definition may depend on f_i and $\mathbf{f}_{\mathcal{N}_i}$; $\chi_i \in \mathbb{R}^n$ is a constant determining the initial state $x_i(0)$, whose value may depend on f_i and $\mathbf{f}_{\mathcal{N}_i}$; $|\cdot|$ denotes the cardinality of a set; and $x_i(t)$, f_i, φ_i, and χ_i are maintained in node i's local memory. The goal of the algorithm (9) and (10) is to steer all the estimates $x_i(t)$'s asymptotically (or, better yet, exponentially) to the unknown x^*, i.e.,

$$\lim_{t \to \infty} x_i(t) = x^*, \quad \forall i \in \mathcal{V}, \tag{11}$$

enabling all the nodes to cooperatively solve problem (8). Note that to realize (9) and (10), for each $i \in \mathcal{V}$, every node $j \in \mathcal{N}_i$ must send node i its $x_j(t)$ at each time t if φ_i does depend on $x_j(t)$, and its f_j at time $t = 0$ if φ_i or χ_i does depend on f_j.

Remark 1. As it turns out and will be shown in Section 4, each φ_i and χ_i in (9) and (10) do not have to depend on $\mathbf{f}_{\mathcal{N}_i}$, so that the nodes do not have to exchange their f_i's. We note that the algorithm PB in [9] also exhibits this feature, but the algorithms PE in [9] and CHE in [10] do not.

4   Zero-Gradient-Sum Algorithms

In this section, we develop a family of algorithms that achieve the stated goal. To facilitate the development, we let $\mathbf{x}^* = (x^*, x^*, \ldots, x^*) \in \mathbb{R}^{nN}$ and $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^{nN}$ denote the minimizer and state vectors, respectively, and write the latter as $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_N(t))$ when we wish to emphasize time or view it as a state trajectory.

Consider a Lyapunov function candidate $V : \mathbb{R}^{nN} \to \mathbb{R}$, defined in terms of the observed f_i's as

$$V(\mathbf{x}) = \sum_{i \in \mathcal{V}} f_i(x^*) - f_i(x_i) - \nabla f_i(x_i)^T (x^* - x_i). \tag{12}$$

4

Notice that V in (12) is continuously differentiable because of Assumption 1, and that it satisfies $V(\mathbf{x}^*) = 0$. Moreover, V is positive definite with respect to $\mathbf{x}^*$ and is radially unbounded, which can be seen by noting that Assumption 1 and the first-order strong convexity condition (2) imply

$$V(\mathbf{x}) \geq \sum_{i \in \mathcal{V}} \frac{\theta_i}{2} \|x^* - x_i\|^2, \quad \forall \mathbf{x} \in \mathbb{R}^{nN}, \tag{13}$$

and (13) in turn implies $V(\mathbf{x}) > 0 \; \forall \mathbf{x} \neq \mathbf{x}^*$ and $V(\mathbf{x}) \to \infty$ as $\|\mathbf{x}\| \to \infty$. Therefore, V in (12) is a legitimate Lyapunov function candidate, which may be used to derive algorithms that ensure (11).

Taking the time derivative of V along the state trajectory $\mathbf{x}(t)$ of the system (9) and calling it $\dot{V} : \mathbb{R}^{nN} \to \mathbb{R}$, we obtain

$$\dot{V}(\mathbf{x}(t)) = \sum_{i \in \mathcal{V}} (x_i(t) - x^*)^T \nabla^2 f_i(x_i(t)) \varphi_i(x_i(t), \mathbf{x}_{\mathcal{N}_i}(t); f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall t \geq 0. \tag{14}$$

Due to Assumption 1 and to each φ_i being locally Lipschitz, $\dot{V}$ in (14) is continuous. In addition, it yields $\dot{V}(\mathbf{x}^*) = 0$. Hence, if the functions $\varphi_i \; \forall i \in \mathcal{V}$ are such that $\dot{V}$ is negative definite with respect to $\mathbf{x}^*$, i.e.,

$$\sum_{i \in \mathcal{V}} (x_i - x^*)^T \nabla^2 f_i(x_i) \varphi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) < 0, \quad \forall \mathbf{x} \neq \mathbf{x}^*, \tag{15}$$

the system (9) would have a unique equilibrium point at $\mathbf{x}^*$, which by the Barbashin-Krasovskii theorem [21] would be globally asymptotically stable. Consequently, regardless of how the constants $\chi_i \; \forall i \in \mathcal{V}$ in (10) are chosen, the goal (11) would be accomplished.

As it follows from the above, the challenge lies in finding $\varphi_i \; \forall i \in \mathcal{V}$, which collectively satisfy (15). Such φ_i's, however, may be difficult to construct because x^* in (15) is unknown to any of the nodes, i.e., x^* depends on *every* f_i via (8), but φ_i maintained by each node $i \in \mathcal{V}$ can *only* depend on f_i and $\mathbf{f}_{\mathcal{N}_i}$. As a result, one cannot let the φ_i's depend on x^*, such as letting $\varphi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = x^* - x_i \; \forall i \in \mathcal{V}$, even though this particular choice guarantees (15) (since each $\nabla^2 f_i(x_i)$ is positive definite, by (4)). Given that the required φ_i's are not readily apparent, instead of searching for them, below we present an alternative approach toward the goal (11), which uses the same V and $\dot{V}$ as in (12) and (14), but demands neither local nor global asymptotic stability.

To state the approach, we first introduce two definitions: let $\mathcal{A} \subset \mathbb{R}^{nN}$ represent the *agreement set* and $\mathcal{M} \subset \mathbb{R}^{nN}$ represent the *zero-gradient-sum manifold*, defined respectively as

$$\mathcal{A} = \{(y_1, y_2, \ldots, y_N) \in \mathbb{R}^{nN} : y_1 = y_2 = \cdots = y_N\}, \tag{16}$$

$$\mathcal{M} = \{(y_1, y_2, \ldots, y_N) \in \mathbb{R}^{nN} : \sum_{i \in \mathcal{V}} \nabla f_i(y_i) = 0\}, \tag{17}$$

so that $\mathbf{x} \in \mathcal{A}$ if and only if all the x_i's agree, and $\mathbf{x} \in \mathcal{M}$ if and only if the sum of all the gradients ∇f_i's, evaluated respectively at the x_i's, is zero. Notice from (16) that $\mathbf{x}^* \in \mathcal{A}$, from (17) and Proposition 1 that $\mathbf{x}^* \in \mathcal{M}$, and from all of them that $\mathbf{x} \in \mathcal{A} \cap \mathcal{M} \Rightarrow \mathbf{x} = \mathbf{x}^*$. Thus, $\mathcal{A} \cap \mathcal{M} = \{\mathbf{x}^*\}$.

Also note from the continuity of each ∇f_i that $\mathcal{M}$ is closed and from the Implicit Function Theorem and the nonsingularity of each $\nabla^2 f_i(x) \ \forall x \in \mathbb{R}^n$ that $\mathcal{M}$ is indeed a manifold of dimension $n(N-1)$.

Having introduced $\mathcal{A}$ and $\mathcal{M}$, we now describe the approach, which is based on the following recognition: to attain the goal (11), condition (15)—which ensures that *every* trajectory $\mathbf{x}(t)$ goes to $\mathbf{x}^*$—is sufficient but not necessary. Rather, all that is needed is a *single* trajectory $\mathbf{x}(t)$, along which $\dot{V}(\mathbf{x}(t)) \leq 0 \ \forall t \geq 0$ and $\lim_{t \to \infty} V(\mathbf{x}(t)) = 0$, since the latter implies (11). Recognizing this, we next derive three conditions on the φ_i's and χ_i's in (9) and (10) that produce such a trajectory. Assume, for a moment, that the χ_i's dictating the initial state $\mathbf{x}(0)$ have been decided, so that we may focus on the φ_i's that shape the trajectory $\mathbf{x}(t)$ leaving $\mathbf{x}(0)$. Observe that $\dot{V}$ in (14) takes the form $\dot{V}(\mathbf{x}(t)) = \Phi_1(\mathbf{x}(t)) - x^{*T}\Phi_2(\mathbf{x}(t)) \ \forall t \geq 0$, where $\Phi_1 : \mathbb{R}^{nN} \to \mathbb{R}$ and $\Phi_2 : \mathbb{R}^{nN} \to \mathbb{R}^n$. Thus, the unknown x^*—which may undesirably affect the sign of $\dot{V}(\mathbf{x}(t))$—can be eliminated by setting $\Phi_2(\mathbf{x}) = 0 \ \forall \mathbf{x} \in \mathbb{R}^{nN}$, i.e., by forcing the φ_i's to satisfy

$$\sum_{i \in \mathcal{V}} \nabla^2 f_i(x_i) \varphi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = 0, \quad \forall \mathbf{x} \in \mathbb{R}^{nN}. \tag{18}$$

With this first condition (18), $\dot{V}$ becomes free of x^*, reducing to

$$\dot{V}(\mathbf{x}(t)) = \sum_{i \in \mathcal{V}} x_i(t)^T \nabla^2 f_i(x_i(t)) \varphi_i(x_i(t), \mathbf{x}_{\mathcal{N}_i}(t); f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall t \geq 0. \tag{19}$$

Next, notice that whenever $\mathbf{x}(t)$ is in the agreement set $\mathcal{A}$, due to (16) and (18), $\dot{V}(\mathbf{x}(t))$ in (19) must vanish. However, whenever $\mathbf{x}(t) \notin \mathcal{A}$, there is no such restriction. Hence, any time $\mathbf{x}(t) \notin \mathcal{A}$, $\dot{V}(\mathbf{x}(t))$ can be made negative by forcing the φ_i's to also satisfy

$$\sum_{i \in \mathcal{V}} x_i^T \nabla^2 f_i(x_i) \varphi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) < 0, \quad \forall \mathbf{x} \in \mathbb{R}^{nN} - \mathcal{A}. \tag{20}$$

With this additional, second condition (20), no matter what x^* is, $\dot{V}(\mathbf{x}(t)) \leq 0$ along $\mathbf{x}(t)$, with equality if and only if $\mathbf{x}(t) \in \mathcal{A}$. Finally, note that (9) and (18) imply

$$\frac{d}{dt} \sum_{i \in \mathcal{V}} \nabla f_i(x_i(t)) = \sum_{i \in \mathcal{V}} \nabla^2 f_i(x_i(t)) \dot{x}_i(t) = 0, \quad \forall t \geq 0,$$

while (11), the continuity of each ∇f_i, and Proposition 1 imply

$$\lim_{t \to \infty} \sum_{i \in \mathcal{V}} \nabla f_i(x_i(t)) = \sum_{i \in \mathcal{V}} \nabla f_i(\lim_{t \to \infty} x_i(t)) = \sum_{i \in \mathcal{V}} \nabla f_i(x^*) = \nabla F(x^*) = 0.$$

The former says that by making the φ_i's satisfy (18), the gradient sum $\sum_{i \in \mathcal{V}} \nabla f_i(x_i(t))$ along $\mathbf{x}(t)$ would remain constant over time, while the latter says that to achieve $\lim_{t \to \infty} V(\mathbf{x}(t)) = 0$ or equivalently (11), this constant sum must be zero, i.e., $\sum_{i \in \mathcal{V}} \nabla f_i(x_i(t)) = 0 \ \forall t \geq 0$. Therefore, in view of (10), the χ_i's must be such that

$$\sum_{i \in \mathcal{V}} \nabla f_i(\chi_i(f_i, \mathbf{f}_{\mathcal{N}_i})) = 0, \tag{21}$$

yielding the third and final condition.

By imposing algebraic constraints on the φ_i's and χ_i's, conditions (18), (20), and (21) characterize a family of algorithms. This family of algorithms share a number of properties, including one that has a nice geometric interpretation: observe from (21), (10), and (17) that $\mathbf{x}(0) \in \mathcal{M}$ and further from (9) and (18) that $\mathbf{x}(t) \in \mathcal{M} \; \forall t > 0$. Thus, every algorithm in the family produces a nonlinear networked dynamical system, whose trajectory $\mathbf{x}(t)$ begins on, and slides along, the zero-gradient-sum manifold $\mathcal{M}$, making $\mathcal{M}$ a positively invariant set. Due to this geometric interpretation, these algorithms are referred to as follows:

Definition 1. A continuous-time distributed algorithm of the form (9) and (10) is said to be a *Zero-Gradient-Sum* (ZGS) algorithm if $\varphi_i \; \forall i \in \mathcal{V}$ are locally Lipschitz and satisfy (18) and (20), and $\chi_i \; \forall i \in \mathcal{V}$ satisfy (21).

The following theorem lists the properties shared by ZGS algorithms, showing that every one of them is capable of asymptotically driving $\mathbf{x}(t)$ to $\mathbf{x}^*$, solving problem (8):

Theorem 1. *Consider the network modeled in Section 3 and the use of a ZGS algorithm described in Definition 1. Suppose Assumption 1 holds. Then: (i) there exists a unique solution $\mathbf{x}(t) \; \forall t \geq 0$ to (9) and (10); (ii) $\mathbf{x}(t) \in \mathcal{M} \; \forall t \geq 0$; (iii) $\dot{V}(\mathbf{x}(t)) \leq 0 \; \forall t \geq 0$, with equality if and only if $\mathbf{x}(t) = \mathbf{x}^*$; (iv) $\lim_{t \to \infty} V(\mathbf{x}(t)) = 0$; and (v) $\lim_{t \to \infty} \mathbf{x}(t) = \mathbf{x}^*$, i.e., (11) holds.*

Proof. Since $\varphi_i \; \forall i \in \mathcal{V}$ are locally Lipschitz, to prove (i) it suffices to show that every solution $\mathbf{x}(t)$ of (9) and (10) lies entirely in a compact subset of $\mathbb{R}^{nN}$. To this end, let $\mathcal{B}(\mathbf{x}^*, r) \subset \mathbb{R}^{nN}$ denote the closed-ball of radius $r \in [0, \infty)$ centered at $\mathbf{x}^*$, i.e., $\mathcal{B}(\mathbf{x}^*, r) = \{\mathbf{y} \in \mathbb{R}^{nN} : \|\mathbf{y} - \mathbf{x}^*\| \leq r\}$. Note from (14), (18), and (20) that $\dot{V}(\mathbf{x}(t)) \leq 0$ along $\mathbf{x}(t)$. This, together with (13), implies that $V(\mathbf{x}(0)) \geq V(\mathbf{x}(t)) \geq \frac{\min_{i \in \mathcal{V}} \theta_i}{2} \|\mathbf{x}(t) - \mathbf{x}^*\|^2$ along $\mathbf{x}(t)$. Hence, $\mathbf{x}(t) \in \mathcal{B}(\mathbf{x}^*, \sqrt{\frac{2V(\mathbf{x}(0))}{\min_{i \in \mathcal{V}} \theta_i}}) \; \forall t \geq 0$, ensuring (i). Statement (ii) has been proven in the paragraph before Definition 1. To verify (iii), notice again from (14), (18), and (20) that $\dot{V}(\mathbf{x}(t)) = 0$ if and only if $\mathbf{x}(t) \in \mathcal{A}$. Due to (ii) and to $\mathcal{A} \cap \mathcal{M} = \{\mathbf{x}^*\}$ shown earlier, (iii) holds. To prove (iv) and (v), we will apply LaSalle's invariance principle from Theorem 4.4 in [21] to the dynamics (9). Let $\Omega = \mathcal{M} \cap \{\mathbf{y} \in \mathbb{R}^{nN} : V(\mathbf{y}) \leq V(\mathbf{x}(0))\}$. Notice that Ω is compact since $\mathcal{M}$ is closed and V in (12) is continuous and satisfies (13). Also note from (17), (9), and (18) that $\mathcal{M}$ is positively invariant, and from (14), (18), (20), $\mathcal{A} \cap \mathcal{M} = \{\mathbf{x}^*\}$, and $\mathbf{x}^* \in \Omega \subset \mathcal{M}$ that $\dot{V}(\mathbf{x}) \leq 0 \; \forall \mathbf{x} \in \Omega$, with equality if and only if $\mathbf{x} = \mathbf{x}^*$. Thus, Ω is positively invariant as well. Moreover, the largest invariant set in $\{\mathbf{y} \in \Omega : \dot{V}(\mathbf{y}) = 0\} = \{\mathbf{x}^*\}$ is $\{\mathbf{x}^*\}$, since it must be nonempty. It follows from Theorem 4.4 in [21] that every solution starting in Ω approaches $\mathbf{x}^*$ as $t \to \infty$, including $\mathbf{x}(t)$. Therefore, (v) holds and, by the continuity of V, (iv) follows. $\square$

Having established Theorem 1, we now present a systematic way to construct ZGS algorithms. First, to find χ_i's that meet condition (21), consider the following proposition, which shows that each f_i has a unique minimizer $x_i^* \in \mathbb{R}^n$:

Proposition 2. *With Assumption 1, for each $i \in \mathcal{V}$, there exists a unique $x_i^* \in \mathbb{R}^n$ such that $f_i(x_i^*) \leq f_i(x) \; \forall x \in \mathbb{R}^n$ and $\nabla f_i(x_i^*) = 0$.*

Proof. See Theorem 6 in [20]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Proposition 2 implies that $\sum_{i \in \mathcal{V}} \nabla f_i(x_i^*) = 0$. Hence, (21) can be met by simply letting

$$\chi_i(f_i, \mathbf{f}_{\mathcal{N}_i}) = x_i^*, \quad \forall i \in \mathcal{V}, \tag{22}$$

which is permissible since every x_i^* in (22) depends just on f_i. It follows that each node $i \in \mathcal{V}$ must solve a "local" convex optimization problem $\min_{x \in \mathbb{R}^n} f_i(x)$ for x_i^* before time $t = 0$, in order to execute (10) and (22). We note, however, that (22) is sufficient for ensuring (21) but not necessary.

Next, to generate locally Lipschitz φ_i's that ensure conditions (18) and (20), notice that each φ_i is premultiplied by $\nabla^2 f_i(x_i)$, which is nonsingular $\forall x_i \in \mathbb{R}^n$. Therefore, the impact of each $\nabla^2 f_i(x_i)$ can be absorbed by setting

$$\varphi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = (\nabla^2 f_i(x_i))^{-1} \phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall i \in \mathcal{V}, \tag{23}$$

where $\phi_i : \mathbb{R}^n \times \mathbb{R}^{n|\mathcal{N}_i|} \to \mathbb{R}^n$ is a locally Lipschitz function of x_i and $\mathbf{x}_{\mathcal{N}_i}$ maintained by node i. For each $i \in \mathcal{V}$, because $\nabla^2 f_i$ is locally Lipschitz (due to Assumption 1) and the determinant of $\nabla^2 f_i(x_i)$ for every $x_i \in \mathbb{R}^n$ is no less than a positive constant θ_i^n (due further to (4)), the mapping $(\nabla^2 f_i(\cdot))^{-1} : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ in (23) is locally Lipschitz. Thus, as long as the ϕ_i's are locally Lipschitz, so would the resulting φ_i's, fulfilling the requirement. With (23), the dynamics (9) become

$$\dot{x}_i(t) = (\nabla^2 f_i(x_i(t)))^{-1} \phi_i(x_i(t), \mathbf{x}_{\mathcal{N}_i}(t); f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall t \geq 0, \; \forall i \in \mathcal{V}, \tag{24}$$

and conditions (18) and (20) simplify to

$$\sum_{i \in \mathcal{V}} \phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = 0, \quad \forall \mathbf{x} \in \mathbb{R}^{nN}, \tag{25}$$

$$\sum_{i \in \mathcal{V}} x_i^T \phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) < 0, \quad \forall \mathbf{x} \in \mathbb{R}^{nN} - \mathcal{A}. \tag{26}$$

Finally, to come up with locally Lipschitz ϕ_i's that assure conditions (25) and (26), suppose each ϕ_i is decomposed as

$$\phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = \sum_{j \in \mathcal{N}_i} \phi_{ij}(x_i, x_j; f_i, f_j), \quad \forall i \in \mathcal{V}, \tag{27}$$

so that the dynamics (24) become

$$\dot{x}_i(t) = (\nabla^2 f_i(x_i(t)))^{-1} \sum_{j \in \mathcal{N}_i} \phi_{ij}(x_i(t), x_j(t); f_i, f_j), \quad \forall t \geq 0, \; \forall i \in \mathcal{V}, \tag{28}$$

8

where $\phi_{ij} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a locally Lipschitz function of x_i and x_j maintained by node i. Then, (25) can be ensured by requiring that every ϕ_{ij} and ϕ_{ji} pair be negative of each other, i.e.,

$$\phi_{ij}(y, z; f_i, f_j) = -\phi_{ji}(z, y; f_j, f_i), \quad \forall i \in \mathcal{V}, \ \forall j \in \mathcal{N}_i, \ \forall y, z \in \mathbb{R}^n, \tag{29}$$

since $\sum_{i \in \mathcal{V}} \phi_i = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \phi_{ij} = \sum_{\{i,j\} \in \mathcal{E}} \phi_{ij} + \phi_{ji} = 0$. With (27) and (29), the left-hand side of (26) turns into

$$\sum_{i \in \mathcal{V}} x_i^T \phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = \frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} (x_i - x_j)^T \phi_{ij}(x_i, x_j; f_i, f_j), \quad \forall \mathbf{x} \in \mathbb{R}^{nN}. \tag{30}$$

Because the graph $\mathcal{G}$ is connected, for any $\mathbf{x} \in \mathbb{R}^{nN} - \mathcal{A}$, there exist $i \in \mathcal{V}$ and $j \in \mathcal{N}_i$ such that $x_i - x_j$ in (30) is nonzero. Hence, (26) can be guaranteed by requiring the ϕ_{ij}'s to also satisfy

$$(y - z)^T \phi_{ij}(y, z; f_i, f_j) < 0, \quad \forall i \in \mathcal{V}, \ \forall j \in \mathcal{N}_i, \ \forall y, z \in \mathbb{R}^n, \ y \neq z. \tag{31}$$

Note that if (29) holds, then ϕ_{ij} satisfies the inequality in (31) if and only if ϕ_{ji} does. Therefore, every pair of neighboring nodes $i, j \in \mathcal{V}$ need only minimal coordination before time $t = 0$ to realize the dynamics (28): only one of them, say, node i, needs to construct a ϕ_{ij} that satisfies the inequality in (31), and the other, i.e., node j, only needs to make sure that $\phi_{ji} = -\phi_{ij}$.

Examples 1 and 2 below illustrate two concrete ways to construct ϕ_{ij}'s that obey (29) and (31):

Example 1. Let $\phi_{ij}(y, z; f_i, f_j) = (\psi_{ij1}(y_1, z_1), \psi_{ij2}(y_2, z_2), \ldots, \psi_{ijn}(y_n, z_n)) \ \forall i \in \mathcal{V} \ \forall j \in \mathcal{N}_i \ \forall y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n \ \forall z = (z_1, z_2, \ldots, z_n) \in \mathbb{R}^n$, where each $\psi_{ij\ell} : \mathbb{R}^2 \to \mathbb{R}$ can be *any* locally Lipschitz function satisfying $\psi_{ij\ell}(y_\ell, z_\ell) = -\psi_{ji\ell}(z_\ell, y_\ell)$ and $(y_\ell - z_\ell)\psi_{ij\ell}(y_\ell, z_\ell) < 0 \ \forall y_\ell \neq z_\ell$ (e.g., $\psi_{ij\ell}(y_\ell, z_\ell) = \tanh(z_\ell - y_\ell)$ or $\psi_{ij\ell}(y_\ell, z_\ell) = -\psi_{ji\ell}(z_\ell, y_\ell) = \frac{z_\ell - y_\ell}{1 + y_\ell^2}$). Then, (29) and (31) hold. ∎

Example 2. Let $\phi_{ij}(y, z; f_i, f_j) = \nabla g_{\{i,j\}}(z) - \nabla g_{\{i,j\}}(y) \ \forall i \in \mathcal{V} \ \forall j \in \mathcal{N}_i \ \forall y, z \in \mathbb{R}^n$, where each $g_{\{i,j\}} : \mathbb{R}^n \to \mathbb{R}$ can be *any* twice continuously differentiable and locally strongly convex function associated with link $\{i, j\} \in \mathcal{E}$ (e.g., $g_{\{i,j\}}(y) = \frac{1}{2} y^T A_{\{i,j\}} y$ where $A_{\{i,j\}} \in \mathbb{R}^{n \times n}$ is any symmetric positive definite matrix, or $g_{\{i,j\}}(y) = f_i(y) + f_j(y)$ if the nodes do not mind exchanging their f_i's). Then, (29) and (31) hold. ∎

Examples 3 and 4 below show that some of the continuous-time distributed consensus algorithms in the literature are special cases of ZGS algorithms. In addition, they are just a slight modification away from solving general unconstrained, separable, convex optimization problems:

Example 3. Consider the scalar (i.e., $n = 1$) linear consensus algorithm $\dot{x}_i(t) = \sum_{j \in \mathcal{N}_i} a_{ij}(x_j(t) - x_i(t)) \ \forall t \geq 0 \ \forall i \in \mathcal{V}$ with symmetric parameters $a_{ij} = a_{ji} > 0 \ \forall \{i, j\} \in \mathcal{E}$ and arbitrary initial states $x_i(0) = y_i \ \forall i \in \mathcal{V}$, studied in [12–14, 16]. By Definition 1 and Theorem 1, this algorithm is a ZGS algorithm that solves problem (8) for $f_i(x) = \frac{1}{2}(x - y_i)^2 \ \forall i \in \mathcal{V}$. Moreover, the algorithm is only a Hessian inverse and an initial condition away (i.e., $\dot{x}_i(t) = (\nabla^2 f_i(x_i(t)))^{-1} \sum_{j \in \mathcal{N}_i} a_{ij}(x_j(t) - x_i(t))$ with $x_i(0) = x_i^*$) from solving *any* convex optimization problem of the form (8) for any $n \geq 1$. Note that the same can be said about the scalar nonlinear consensus protocol in [11]. ∎

9

Example 4. Consider the multivariable (i.e., $n \geq 1$) weighted-average consensus algorithm $\dot{x}_i(t) = W_i^{-1} \sum_{j \in \mathcal{N}_i} (x_j(t) - x_i(t)) \; \forall t \geq 0 \; \forall i \in \mathcal{V}$ with $W_i = W_i^T > 0$ and $x_i(0) = y_i$, proposed in [15] as a step toward a distributed Kalman filter. This algorithm is a ZGS algorithm that solves problem (8) for $f_i(x) = \frac{1}{2}(x - y_i)^T W_i (x - y_i) \; \forall i \in \mathcal{V}$. Indeed, it is only a replacement of W_i^{-1} by $(\nabla^2 f_i(x_i(t)))^{-1}$ and $x_i(0) = y_i$ by $x_i(0) = x_i^*$ away from solving for general f_i's. ∎

5   Convergence Rate Analysis

In this section, we derive lower and upper bounds on the exponential convergence rates of the ZGS algorithms described in (28) and Example 2, i.e.,

$$\dot{x}_i(t) = (\nabla^2 f_i(x_i(t)))^{-1} \sum_{j \in \mathcal{N}_i} \nabla g_{\{i,j\}}(x_j(t)) - \nabla g_{\{i,j\}}(x_i(t)), \quad \forall t \geq 0, \; \forall i \in \mathcal{V}, \tag{32}$$

which form a subset of those in Definition 1, but include the ones in Examples 3 and 4 as a subset. To enable the derivation, suppose an initial state $\mathbf{x}(0) \in \mathcal{M}$ is given (e.g., $\mathbf{x}(0) = (x_1^*, x_2^*, \ldots, x_N^*)$ as in (10) and (22)). With this $\mathbf{x}(0)$, let $\mathcal{C}_i = \{x \in \mathbb{R}^n : f_i(x^*) - f_i(x) - \nabla f_i(x)^T (x^* - x) \leq V(\mathbf{x}(0))\}$ $\forall i \in \mathcal{V}$ and let $\mathcal{C} = \mathrm{conv} \cup_{i \in \mathcal{V}} \mathcal{C}_i$, where conv denotes the convex hull. It follows from Assumption 1, (2), (12), and (iii) in Theorem 1 that $\mathcal{C}_i \; \forall i \in \mathcal{V}$ are compact, $\mathcal{C}$ is convex and compact, and

$$x_i(t), x^* \in \mathcal{C}_i \subset \mathcal{C}, \quad \forall t \geq 0, \; \forall i \in \mathcal{V}. \tag{33}$$

For each $i \in \mathcal{V}$, due to Assumption 1, (4), and $\mathcal{C}$ being compact, there exists a $\Theta_i \geq \theta_i$ such that

$$\nabla^2 f_i(x) \leq \Theta_i I_n, \quad \forall x \in \mathcal{C}. \tag{34}$$

Moreover, for each $\{i, j\} \in \mathcal{E}$, due to (3), $g_{\{i,j\}}$ being locally strongly convex, and $\mathcal{C}$ being convex and compact, there exists a $\gamma_{\{i,j\}} > 0$ such that

$$(\nabla g_{\{i,j\}}(y) - \nabla g_{\{i,j\}}(x))^T (y - x) \geq \gamma_{\{i,j\}} \|y - x\|^2, \quad \forall x, y \in \mathcal{C}. \tag{35}$$

Furthermore, for each $\{i, j\} \in \mathcal{E}$, due to (3), (4), (35), $\nabla^2 g_{\{i,j\}}$ being continuous, and $\mathcal{C}$ being convex and compact, there exists a $\Gamma_{\{i,j\}} \geq \gamma_{\{i,j\}}$ such that

$$\nabla^2 g_{\{i,j\}}(x) \leq \Gamma_{\{i,j\}} I_n, \quad \forall x \in \mathcal{C}. \tag{36}$$

Observe that the constants Θ_i's, $\gamma_{\{i,j\}}$'s, and $\Gamma_{\{i,j\}}$'s—unlike the convexity parameters θ_i's—depend on the initial state $\mathbf{x}(0)$ via the sets $\mathcal{C}$ and $\mathcal{C}_i$'s. Thus, the convergence rate results obtained below are dependent on $\mathbf{x}(0)$ in general. One exception is the case where the f_i's and $g_{\{i,j\}}$'s are quadratic functions, for which the θ_i's, Θ_i's, $\gamma_{\{i,j\}}$'s, and $\Gamma_{\{i,j\}}$'s may be taken as the smallest and largest eigenvalues of the Hessians of the f_i's and $g_{\{i,j\}}$'s, respectively, independent of $\mathbf{x}(0)$. Finally, for convenience, let $\theta = \min_{i \in \mathcal{V}} \theta_i$, $\Theta = \max_{i \in \mathcal{V}} \Theta_i$, $\gamma = \min_{\{i,j\} \in \mathcal{E}} \gamma_{\{i,j\}}$, and $\Gamma = \max_{\{i,j\} \in \mathcal{E}} \Gamma_{\{i,j\}}$.

The following theorem establishes the exponential convergence of the ZGS algorithms (32) and provides a lower bound ρ on their convergence rates, that they can do no worse than:

Theorem 2. *Consider the network modeled in Section 3 and the use of a ZGS algorithm described in (32). Suppose Assumption 1 holds. Then,*

$$V(\mathbf{x}(t)) \leq V(\mathbf{x}(0))e^{-\rho t}, \quad \forall t \geq 0, \tag{37}$$

$$\sum_{i \in \mathcal{V}} \theta_i \|x_i(t) - x^*\|^2 \leq \sum_{i \in \mathcal{V}} \Theta_i \|x_i(0) - x^*\|^2 e^{-\rho t}, \quad \forall t \geq 0, \tag{38}$$

where $\rho = \sup\{\varepsilon \in \mathbb{R} : \varepsilon P \leq Q\} > 0$, $P = [P_{ij}] \in \mathbb{R}^{N \times N}$ *is a positive semidefinite matrix given by*

$$P_{ij} = \begin{cases} (\frac{1}{2} - \frac{1}{N})\Theta_i + \frac{1}{2N^2}\sum_{\ell \in \mathcal{V}} \Theta_\ell, & \text{if } i = j, \\ -\frac{\Theta_i + \Theta_j}{2N} + \frac{1}{2N^2}\sum_{\ell \in \mathcal{V}} \Theta_\ell, & \text{otherwise}, \end{cases} \tag{39}$$

and $Q = [Q_{ij}] \in \mathbb{R}^{N \times N}$ *is a positive semidefinite matrix given by*

$$Q_{ij} = \begin{cases} \sum_{\ell \in \mathcal{N}_i} \gamma_{\{i,\ell\}}, & \text{if } i = j, \\ -\gamma_{\{i,j\}}, & \text{if } \{i,j\} \in \mathcal{E}, \\ 0, & \text{otherwise}. \end{cases} \tag{40}$$

Proof. Let $\eta(t) = \frac{1}{N}\sum_{j \in \mathcal{V}} x_j(t) \ \forall t \geq 0$. Due to (33) and the convexity of $\mathcal{C}$, $\eta(t) \in \mathcal{C}$. Moreover, by Proposition 1, $\sum_{i \in \mathcal{V}} f_i(x^*) = F(x^*) \leq F(\eta(t)) = \sum_{i \in \mathcal{V}} f_i(\eta(t))$. Observe from (17) and (ii) in Theorem 1 that $\sum_{i \in \mathcal{V}} \nabla f_i(x_i(t)) = 0$. Thus, from (12), $V(\mathbf{x}(t)) \leq \sum_{i \in \mathcal{V}} f_i(\eta(t)) - f_i(x_i(t)) - \nabla f_i(x_i(t))^T(\eta(t) - x_i(t))$. It follows from (5), (7), (34), (33), and (39) that

$$V(\mathbf{x}(t)) \leq \sum_{i \in \mathcal{V}} \frac{\Theta_i}{2}\|x_i(t) - \frac{1}{N}\sum_{j \in \mathcal{V}} x_j(t)\|^2 = \mathbf{x}(t)^T(P \otimes I_n)\mathbf{x}(t), \quad \forall t \geq 0, \tag{41}$$

where $\otimes$ denotes the Kronecker product. Next, using (32), (9), and (19), we can write

$$\dot{V}(\mathbf{x}(t)) = -\frac{1}{2}\sum_{i \in \mathcal{V}}\sum_{j \in \mathcal{N}_i}(x_j(t) - x_i(t))^T(\nabla g_{\{i,j\}}(x_j(t)) - \nabla g_{\{i,j\}}(x_i(t))), \quad \forall t \geq 0. \tag{42}$$

Therefore, from (35), (33), and (40),

$$-\dot{V}(\mathbf{x}(t)) \geq \frac{1}{2}\sum_{i \in \mathcal{V}}\sum_{j \in \mathcal{N}_i} \gamma_{\{i,j\}}\|x_j(t) - x_i(t)\|^2 = \mathbf{x}(t)^T(Q \otimes I_n)\mathbf{x}(t), \quad \forall t \geq 0. \tag{43}$$

To relate (41) and (43), notice from (39) and (40) that both P and Q are symmetric with zero row sums. Also, $\forall y = (y_1, y_2, \ldots, y_N) \in \mathbb{R}^N$, $y^T P y = \sum_{i \in \mathcal{V}} \frac{\Theta_i}{2}(y_i - \frac{1}{N}\sum_{j \in \mathcal{V}} y_j)^2 \geq 0$ and $y^T Q y = \frac{1}{2}\sum_{i \in \mathcal{V}}\sum_{j \in \mathcal{N}_i} \gamma_{\{i,j\}}(y_j - y_i)^2 \geq 0$, where the equalities hold if and only if $y_1 = y_2 = \cdots = y_N$. Hence, both P and Q are positive semidefinite with $N - 1$ positive eigenvalues, one eigenvalue at 0, and $(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}})$ being its corresponding eigenvector. It follows that there exists an orthogonal $W \in \mathbb{R}^{N \times N}$ with the first column being $(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}})$, such that $W^T P W = \text{diag}(0, \bar{P})$ and $W^T Q W = \text{diag}(0, \bar{Q})$, where $\bar{P}, \bar{Q} \in \mathbb{R}^{(N-1) \times (N-1)}$, $\bar{P} = \bar{P}^T > 0$, and $\bar{Q} = \bar{Q}^T > 0$. Note that

$\forall \varepsilon \in \mathbb{R}$, $\varepsilon P \leq Q \Leftrightarrow \varepsilon \bar{P} \leq \bar{Q} \Leftrightarrow \varepsilon I_{N-1} \leq \bar{P}^{-1/2}\bar{Q}\bar{P}^{-1/2}$, where $\bar{P}^{1/2} = (\bar{P}^{1/2})^T > 0$ is the square root of $\bar{P}$ via the spectral decomposition, i.e., $\bar{P} = \bar{P}^{1/2}\bar{P}^{1/2}$. Since $\rho = \sup\{\varepsilon \in \mathbb{R} : \varepsilon P \leq Q\}$ and $\bar{P}^{-1/2}\bar{Q}\bar{P}^{-1/2} = (\bar{P}^{-1/2}\bar{Q}\bar{P}^{-1/2})^T > 0$, ρ is the smallest eigenvalue of $\bar{P}^{-1/2}\bar{Q}\bar{P}^{-1/2}$ which is positive and satisfies $\rho P \leq Q$. Therefore, $\rho(P \otimes I_n) \leq Q \otimes I_n$. This, along with (41) and (43), implies $\rho V(\mathbf{x}(t)) \leq -\dot{V}(\mathbf{x}(t))$, i.e., (37). Finally, due to (2), (12), (37), (34), (7), (5), and (33), $\sum_{i \in \mathcal{V}} \frac{\theta_i}{2}\|x_i(t) - x^*\|^2 \leq V(\mathbf{x}(t)) \leq V(\mathbf{x}(0))e^{-\rho t} \leq \sum_{i \in \mathcal{V}} \frac{\Theta_i}{2}\|x_i(0) - x^*\|^2 e^{-\rho t}$, i.e., (38) holds. $\quad\square$

The lower bound ρ in Theorem 2 can be calculated according to its proof: ρ is the smallest eigenvalue of $\bar{P}^{-1/2}\bar{Q}\bar{P}^{-1/2}$. The corollary below gives another lower bound, which is not as tight as ρ but is explicit in the algebraic connectivity $\lambda_2 > 0$ of the graph $\mathcal{G}$:

Corollary 1. *With the setup of Theorem 2,*

$$V(\mathbf{x}(t)) \leq V(\mathbf{x}(0))e^{-\frac{2\gamma}{\Theta}\lambda_2 t}, \quad \forall t \geq 0, \tag{44}$$

$$\|\mathbf{x}(t) - \mathbf{x}^*\| \leq \sqrt{\frac{\Theta}{\theta}}\|\mathbf{x}(0) - \mathbf{x}^*\|e^{-\frac{\gamma}{\Theta}\lambda_2 t}, \quad \forall t \geq 0. \tag{45}$$

Proof. From (41) and (43),

$$V(\mathbf{x}(t)) \leq \frac{\Theta}{2}\sum_{i \in \mathcal{V}}\|x_i(t) - \frac{1}{N}\sum_{j \in \mathcal{V}} x_j(t)\|^2 = \frac{\Theta}{2N}\mathbf{x}(t)^T(\mathcal{L}_{\bar{\mathcal{G}}} \otimes I_n)\mathbf{x}(t), \quad \forall t \geq 0, \tag{46}$$

$$-\dot{V}(\mathbf{x}(t)) \geq \frac{\gamma}{2}\sum_{i \in \mathcal{V}}\sum_{j \in \mathcal{N}_i}\|x_j(t) - x_i(t)\|^2 = \gamma\mathbf{x}(t)^T(\mathcal{L}_{\mathcal{G}} \otimes I_n)\mathbf{x}(t), \quad \forall t \geq 0, \tag{47}$$

where $\mathcal{L}_{\bar{\mathcal{G}}} \in \mathbb{R}^{N \times N}$ is the Laplacian of the complete graph $\bar{\mathcal{G}}$ with vertex set $\mathcal{V}$, and $\mathcal{L}_{\mathcal{G}} \in \mathbb{R}^{N \times N}$ is the Laplacian of $\mathcal{G}$. Obviously, $\mathcal{L}_{\bar{\mathcal{G}}}$ has $N-1$ eigenvalues at N, $\mathcal{L}_{\mathcal{G}}$ has $N-1$ positive eigenvalues among which λ_2 is the smallest, and both $\mathcal{L}_{\bar{\mathcal{G}}}$ and $\mathcal{L}_{\mathcal{G}}$ have one eigenvalue at 0 with $(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}})$ being its eigenvector. Let $W \in \mathbb{R}^{N \times N}$ contain N orthonormal eigenvectors of $\mathcal{L}_{\mathcal{G}}$ in its columns. Then, $W^T\mathcal{L}_{\bar{\mathcal{G}}}W$ and $W^T\mathcal{L}_{\mathcal{G}}W$ are diagonal matrices similar to $\mathcal{L}_{\bar{\mathcal{G}}}$ and $\mathcal{L}_{\mathcal{G}}$, and both contain the eigenvalue 0 in the same diagonal position. Hence, $\lambda_2 W^T\mathcal{L}_{\bar{\mathcal{G}}}W \leq NW^T\mathcal{L}_{\mathcal{G}}W$, so that $\lambda_2\mathcal{L}_{\bar{\mathcal{G}}} \leq N\mathcal{L}_{\mathcal{G}}$. Applying this inequality to (46) and (47), we get $\frac{2\gamma}{\Theta}\lambda_2 V(\mathbf{x}(t)) \leq -\dot{V}(\mathbf{x}(t))$, i.e., (44). Finally, (45) follows from (44) the same way (38) does from (37). $\quad\square$

Notice that in the special case where $n = 1$, $f_i(x) = \frac{1}{2}(x - x_i^*)^2$ $\forall i \in \mathcal{V}$, and $g_{\{i,j\}}(x) = \frac{1}{2}x^2$ $\forall\{i,j\} \in \mathcal{E}$, we may let the θ_i's, Θ_i's, and $\gamma_{\{i,j\}}$'s all be 1. In this case, Theorem 2 and Corollary 1 both yield $\|\mathbf{x}(t) - \mathbf{x}^*\| \leq \|\mathbf{x}(0) - \mathbf{x}^*\|e^{-\lambda_2 t}$ $\forall t \geq 0$, which coincides with the well-known convergence rate result for the linear consensus algorithm $\dot{x}_i(t) = \sum_{j \in \mathcal{N}_i} x_j(t) - x_i(t)$ $\forall t \geq 0$ $\forall i \in \mathcal{V}$, reported in [12]. Hence, Theorem 2 and Corollary 1 may be regarded as a generalization of such a result for distributed consensus, to distributed convex optimization.

The next theorem looks at the performance of the ZGS algorithms (32) from the other end, providing an upper bound $\tilde{\rho}$ on their exponential convergence rates that mirrors Theorem 2:

Theorem 3. *Consider the network modeled in Section 3 and the use of a ZGS algorithm described in (32). Suppose Assumption 1 holds. Then,*

$$V(\mathbf{x}(t)) \geq V(\mathbf{x}(0))e^{-\tilde{\rho}t}, \quad \forall t \geq 0, \tag{48}$$

$$\sum_{i\in\mathcal{V}} \Theta_i \|x_i(t) - x^*\|^2 \geq \sum_{i\in\mathcal{V}} \theta_i \|x_i(0) - x^*\|^2 e^{-\tilde{\rho}t}, \quad \forall t \geq 0, \tag{49}$$

where $\tilde{\rho} = \inf\{\varepsilon \in \mathbb{R} : \varepsilon\tilde{P} \geq \tilde{Q}\} > 0$, $\tilde{P} \in \mathbb{R}^{N\times N}$ is a positive definite matrix given by $\tilde{P} = \mathrm{diag}(\frac{\theta_1}{2}, \frac{\theta_2}{2}, \ldots, \frac{\theta_N}{2})$, and $\tilde{Q} = [\tilde{Q}_{ij}] \in \mathbb{R}^{N\times N}$ is a positive semidefinite matrix given by

$$\tilde{Q}_{ij} = \begin{cases} \sum_{\ell\in\mathcal{N}_i} \Gamma_{\{i,\ell\}}, & \text{if } i = j, \\ -\Gamma_{\{i,j\}}, & \text{if } \{i,j\} \in \mathcal{E}, \\ 0, & \text{otherwise.} \end{cases} \tag{50}$$

Proof. From (2) and (12), $V(\mathbf{x}(t)) \geq \sum_{i\in\mathcal{V}} \frac{\theta_i}{2} \|x_i(t) - x^*\|^2 = (\mathbf{x}(t) - \mathbf{x}^*)^T(\tilde{P} \otimes I_n)(\mathbf{x}(t) - \mathbf{x}^*)$ $\forall t \geq 0$. From (42), (36), (7), (6), (33), and (50), $-\dot{V}(\mathbf{x}(t)) \leq \frac{1}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i} \Gamma_{\{i,j\}} \|x_j(t) - x_i(t)\|^2 = \frac{1}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i} \Gamma_{\{i,j\}} \|(x_j(t) - x^*) - (x_i(t) - x^*)\|^2 = (\mathbf{x}(t) - \mathbf{x}^*)^T(\tilde{Q} \otimes I_n)(\mathbf{x}(t) - \mathbf{x}^*) \ \forall t \geq 0$. Like Q in (40), $\tilde{Q}$ in (50) is symmetric positive semidefinite with exactly one eigenvalue at 0. Thus, so is $\tilde{P}^{-1/2}\tilde{Q}\tilde{P}^{-1/2}$, where $\tilde{P}^{1/2} = \mathrm{diag}(\sqrt{\frac{\theta_1}{2}}, \sqrt{\frac{\theta_2}{2}}, \ldots, \sqrt{\frac{\theta_N}{2}})$ is the square root of $\tilde{P}$. Since $\tilde{\rho} = \inf\{\varepsilon \in \mathbb{R} : \varepsilon\tilde{P} \geq \tilde{Q}\}$ and $\forall\varepsilon \in \mathbb{R}$, $\varepsilon\tilde{P} \geq \tilde{Q} \Leftrightarrow \varepsilon I_N \geq \tilde{P}^{-1/2}\tilde{Q}\tilde{P}^{-1/2}$, $\tilde{\rho}$ is the largest eigenvalue of $\tilde{P}^{-1/2}\tilde{Q}\tilde{P}^{-1/2}$ which is positive and such that $\tilde{\rho}\tilde{P} \geq \tilde{Q}$. Therefore, $\tilde{\rho}V(\mathbf{x}(t)) \geq -\dot{V}(\mathbf{x}(t))$, proving (48). Finally, from (12), (34), (7), (5), (33), (48), and (2), we get (49). $\qquad\square$

In contrast to ρ, the upper bound $\tilde{\rho}$ in Theorem 3 is the largest eigenvalue of $\tilde{P}^{-1/2}\tilde{Q}\tilde{P}^{-1/2}$. The next corollary is to Theorem 3 as Corollary 1 is to Theorem 2, giving another upper bound that is not as tight as $\tilde{\rho}$ but is explicit in the spectral radius $\lambda_N > 0$ of the graph Laplacian $\mathcal{L}_\mathcal{G}$:

Corollary 2. *With the setup of Theorem 3,*

$$V(\mathbf{x}(t)) \geq V(\mathbf{x}(0))e^{-\frac{2\Gamma}{\theta}\lambda_N t}, \quad \forall t \geq 0, \tag{51}$$

$$\|\mathbf{x}(t) - \mathbf{x}^*\| \geq \sqrt{\frac{\theta}{\Theta}} \|\mathbf{x}(0) - \mathbf{x}^*\| e^{-\frac{\Gamma}{\theta}\lambda_N t}, \quad \forall t \geq 0. \tag{52}$$

Proof. From the proof of Theorem 3, $\forall t \geq 0$, we have $V(\mathbf{x}(t)) \geq \sum_{i\in\mathcal{V}} \frac{\theta}{2}\|x_i(t) - x^*\|^2 = \frac{\theta}{2}\|\mathbf{x}(t) - \mathbf{x}^*\|^2$ and $-\dot{V}(\mathbf{x}(t)) \leq \frac{1}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i} \Gamma\|(x_j(t) - x^*) - (x_i(t) - x^*)\|^2 = \Gamma(\mathbf{x}(t) - \mathbf{x}^*)^T(\mathcal{L}_\mathcal{G} \otimes I_n)(\mathbf{x}(t) - \mathbf{x}^*) \leq \Gamma\lambda_N\|\mathbf{x}(t) - \mathbf{x}^*\|^2$. Consequently, $\frac{2\Gamma}{\theta}\lambda_N V(\mathbf{x}(t)) \geq -\dot{V}(\mathbf{x}(t))$, implying that (51) and (52) hold. $\qquad\square$

Note that for the special case below Corollary 1, we may let the $\Gamma_{\{i,j\}}$'s be 1, so that Theorem 3 and Corollary 2 both lead to $\|\mathbf{x}(t) - \mathbf{x}^*\| \geq \|\mathbf{x}(0) - \mathbf{x}^*\|e^{-\lambda_N t} \ \forall t \geq 0$, which is again known. Finally, note that the above analysis provides a framework for studying the interplay among network topologies (i.e., $\mathcal{V}$ and $\mathcal{E}$), problem characteristics (i.e., the f_i's, θ_i's, and Θ_i's), and ZGS algorithm parameters (i.e., the $g_{\{i,j\}}$'s, $\gamma_{\{i,j\}}$'s, and $\Gamma_{\{i,j\}}$'s), which may be worthy of further research.

13

6   Conclusion

In this paper, using a convexity-based Lyapunov function candidate, we have developed a set of continuous-time ZGS algorithms, which solve a class of distributed convex optimization problems over networks. We have established the asymptotic and exponential convergence of these algorithms and derived lower and upper bounds on their convergence rates. We have also shown that the ZGS algorithms for distributed convex optimization are closely related to the basic algorithms for distributed consensus, suggesting that the former may be extended in a number of directions just like the latter were, in ways that possibly parallel the latter.

References

[1] M. G. Rabbat and R. D. Nowak, "Distributed optimization in sensor networks," in *Proc. International Symposium on Information Processing in Sensor Networks*, Berkeley, CA, 2004, pp. 20–27.

[2] D. Blatt, A. O. Hero, and H. Gauchman, "A convergent incremental gradient method with a constant step size," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 29–51, 2007.

[3] B. Johansson, M. Rabi, and M. Johansson, "A simple peer-to-peer algorithm for distributed optimization in sensor networks," in *Proc. IEEE Conference on Decision and Control*, New Orleans, LA, 2007, pp. 4705–4710.

[4] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Incremental stochastic subgradient algorithms for convex optimization," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 691–717, 2009.

[5] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Proc. IEEE Conference on Decision and Control*, Cancun, Mexico, 2008, pp. 4185–4190.

[6] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[7] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[8] ——, "Asynchronous gossip algorithms for stochastic optimization," in *Proc. IEEE Conference on Decision and Control*, Shanghai, China, 2009, pp. 3581–3586.

[9] J. Lu, C. Y. Tang, P. R. Regier, and T. D. Bow, "Gossip algorithms for convex consensus optimization over networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 12, 2011, to appear.

[10] J. Lu, P. R. Regier, and C. Y. Tang, "Control of distributed convex optimization," in *Proc. IEEE Conference on Decision and Control*, Atlanta, GA, 2010, pp. 489–495.

[11] R. Olfati-Saber and R. M. Murray, "Consensus protocols for networks of dynamic agents," in *Proc. American Control Conference*, Denver, CO, 2003, pp. 951–956.

[12] ——, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[13] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1867–1872, 2005.

[14] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 655–661, 2005.

[15] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Distributed sensor fusion using dynamic consensus," in *Proc. IFAC World Congress*, Prague, Czech Republic, 2005.

[16] A. Tahbaz-Salehi and A. Jadbabaie, "Small world phenomenon, rapidly mixing Markov chains, and average consensus algorithms," in *Proc. IEEE Conference on Decision and Control*, New Orleans, LA, 2007, pp. 276–281.

[17] R. Goebel and R. T. Rockafellar, "Local strong convexity and local lipschitz continuity of the gradient of convex functions," *Journal of Convex Analysis*, vol. 15, no. 2, pp. 263–270, 2008.

[18] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course.* Norwell, MA: Kluwer Academic Publishers, 2004.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization.* New York, NY: Cambridge University Press, 2004.

[20] M. S. Bazaraa and J. J. Goode, "On symmetric duality in nonlinear programming," *Operations Research*, vol. 21, no. 1, pp. 1–9, 1973.

[21] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.